Navy Experimental Diving Unit
321 Bullfinch Rd.
Panama City, FL 32407-7015

NEDU TR 12-03
October 2012

# THE USE OF ONE-SAMPLE PREDICTION INTERVALS FOR ESTIMATING $CO_2$ SCRUBBER CANISTER DURATIONS

NAVSEA
NAVAL SEA SYSTEMS COMMAND
Navy Experimental Diving Unit

Authors: J. R. CLARKE
V. FERRIS

| REPORT DOCUMENTATION PAGE | *Form Approved* *OMB No. 0704-0188* |
|---|---|



| | | |
|---|---|---|
| **1. REPORT DATE** *(DD-MM-YYYY)* **Nov 6, 2012** | **2. REPORT TYPE** **Tech Report** | **3. DATES COVERED** *(From - To)* July 2012-Nov2012 |

| | |
|---|---|
| **4. TITLE AND SUBTITLE** The Use Of One-Sample Prediction Intervals For Estimating $Co_2$ Scrubber Canister Durations | **5a. CONTRACT NUMBER** |
| | **5b. GRANT NUMBER** |
| | **5c. PROGRAM ELEMENT NUMBER** |
| **6. AUTHOR(S)** Dr. John Clarke and Vince Ferris | **5d. PROJECT NUMBER** **TR 12-03** |
| | **5e. TASK NUMBER** **LT 09-01** |
| | **5f. WORK UNIT NUMBER** |
| **7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)** Navy Experimental Diving Unit<br>321 Bullfinch Rd.<br>Panama City, FL 32407 | **8. PERFORMING ORGANIZATION REPORT NUMBER** |
| **9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)** Naval Sea Systems Command<br>1333 Isaac Hull Ave., SE<br>Washington Navy Yard, DC 20376 | **10. SPONSOR/MONITOR'S ACRONYM(S)** |
| | **11. SPONSOR/MONITOR'S REPORT NUMBER(S)** |

**12. DISTRIBUTION / AVAILABILITY STATEMENT**
Distribution Statement A:
Approved for public release; distribution is unlimited.

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT**
One-sample prediction intervals can be used to estimate statistically safe duration bounds for rebreather carbon dioxide scrubber canisters in the absence of complete temperature characterization. Examples from the academic literature are shown, and results specific to actual canister durations are compared using Monte Carlo simulation and inverse cumulative distribution functions for the Gaussian distribution.

**15. SUBJECT TERMS**
tests of normality, prediction interval, confidence interval, $CO_2$ scrubber, canister bounds, Monte Carlo simulation, inverse cumulative distribution function

| **16. SECURITY CLASSIFICATION OF:** | | | **17. LIMITATION OF ABSTRACT** | **18. NUMBER OF PAGES** | **19a. NAME OF RESPONSIBLE PERSON** Nancy Hicks |
|---|---|---|---|---|---|
| **a. REPORT** (U) | **b. ABSTRACT** (U) | **c. THIS PAGE** (U) | | 23 | **19b. TELEPHONE NUMBER** *(include area code)* 850-230-3170 |

**Standard Form 298 (Rev. 8-98)**
**Prescribed by ANSI Std. Z39.18**

## CONTENTS



**APPENDICES:**

# INTRODUCTION

Navy Experimental Diving Unit (NEDU) Technical Report 2-99[1] describes the rationale and procedure for using canister breakthrough prediction limits for rebreathers in Authorized for Navy Use (ANU) testing, which requires underwater breathing apparatus (UBAs) to be tested at multiple temperatures over the whole range of operationally relevant water temperatures. Appendix A is excerpted from that report. The current report, a follow-on to NEDU TR 2-99, specifically continues NEDU's explanation of the use of small samples to establish prediction interval bounds (in the special case where only one water temperature is used for testing) that describe the characteristics of carbon dioxide absorbent canisters in closed- and semiclosed-circuit UBAs (rebreathers). The described method is useful for estimating how long a diver can safely use a rebreather during "quick-look" evaluations of new rebreathers. Quick or "first-look" evaluations are designed to screen UBAs before the complete testing required for giving AMU (Approved for Military Use) status to a UBA.

# BACKGROUND

NEDU first used regression techniques and confidence intervals to develop canister limits in its report TR 2-93, *MK 16 Canister Limits for SDV Operations.*[2] Prediction intervals first appeared in NEDU TR 09-97*, Recommended Canister Limits for the Draeger LAR V/MK 25 UBA Using 408 L-Grade and 812 D-Grade Sofnolime.*[3]

<u>Definitions</u>

According to Devore,[4]

> A CI (confidence interval) refers to a parameter, or population characteristic, whose value is fixed but unknown to us. In contrast, a <u>future</u> value of Y is not a parameter but instead a random variable; for this reason we refer to an interval of plausible values for a future Y as a **prediction interval** rather than a confidence interval.

In keeping with standard statistical nomenclature, we refer to a "sample" as a set of observations or individuals from a parent population. That parent population is usually referred to as simply the "population," and we assume that its true characteristics are unknowable. However, we can infer its characteristics from the statistical properties of the sample. Sample sizes may range from two to hundreds, but in diving studies they are typically smaller than fifty. For unmanned tests of diving equipment, a sample size of five is usually used.

<u>One-Sample Prediction Limits</u>

In the statistical literature, prediction limits are most often discussed with reference to regression. Indeed, that is the manner in which such limits are used in studies of canister duration during AMU testing. As described in TR 2-99,[1] canister durations are temperature dependent, and thus complete canister studies use four or five test temperatures in characterizing the canister. Each test at each temperature is considered a separate sample.

Each test temperature is a preselected value in the domain of the predictor variable (independent variable), *x* and the resulting correlated canister durations in minutes, are values for the response variable (dependent variable), y. Predicted values for the response variable for values within the domain of the predictor variable are expressed using ŷ.

Prediction limits can also be used when canister duration data is acquired at only one temperature. Appendix A includes examples of such usage in the engineering statistics literature[4-6]. For this report, since the data used was collected at a single temperature and regression analysis was therefore not used, the phrase *prediction interval bounds* more accurately describes the predicted canister duration, and will therefore be used exclusively.

## PROCEDURES

To illustrate the methodology of one-sample prediction intervals, a textbook example is found on page 299 of Devore.[1] In a journal article on the texture of frankfurters,[7] the following percentages of fat were found in a sample of 10 frankfurters: 25.2, 21.3, 22.8, 17.0, 29.8, 21.0, 25.5, 16.0, 20.9, and 19.5%. The mean value of the percentage of fat for this data was 21.90%, and the standard deviation was 4.13%.

### Tests for Normality

In principle, before applying confidence or prediction intervals to data, the data should be tested for normality. One of the best known tests for normality is the Kolmogorov-Smirnov (K-S) normality test. In the frankfurter example, the Kolmogorov-Smirnov normality test yielded a K-S distance of 0.158 and an associated $p = 0.601$. For the K-S test, this is a passing value, meaning that the sample data seems to match the pattern to be expected if the data had been drawn from a population with a normal distribution.

However, normality testing entails several caveats, so the positive results are not as reassuring as one might think. First, the K-S test is not properly applied without the Dallal-Wilkinson-Lilliefors corrections, and not all commercial software uses those corrections. Secondly, and more importantly, small sample sizes almost always pass a normality test. So the test has little power to help the researcher decide whether parametric tests or nonparametric tests, are more appropriate.

Tests of normality that are reputed to be better than the corrected K-S test are the Shapiro-Wilk normality test or the D'Agostino-Pearson omnibus test, a more generally useful test. Due to the complexity of normality test interpretation, statisticians in human clinical trials often use a suite of tests to make their normality decisions. An example in the anesthesiology literature[8] analyzed experimental data thusly: the K-S, the Martinez-Iglewicz, and the D'Agostino-Pearson Omnibus $K^2$ tests were used to test continuous and numerical data for normality. Skewed data were summarized as median (range) and analyzed for between-group differences via Kruskal-Wallis tests. Exactly how these statistical test results are used in the decision-making process is seldom, if ever, described in the literature.

For our purposes, if small sample sizes result in data so skewed that normality tests — K-S or otherwise — are not passed, then the researcher should look for sample preparation or experimental problems. The data, whether it be of fat concentrations or canister durations, should not be used to make predictions of anything.

Confidence Intervals

Once it has been demonstrated that the normality assumption of the data is reasonable, then the confidence interval (C.I.) on the mean for the data can be calculated for a given confidence level. A confidence level percentile is defined as 100(1-α)%, where the significance level is α and confidence coefficient is (1-α). The confidence coefficient is the proportion of times that the population is sampled that the C.I. actually does contain the population parameter of interest. Typical confidence level percentiles used to express a C.I. are 90%, 95% and 99%. For a selected confidence level percentile, the C.I. lower- and upper-confidence limits for the population mean, μ, are calculated using equation (1).

$$\left[\mu_{lower}, \mu_{upper}\right] = \bar{x} \pm t_{(1-\alpha/2;\, n-1)} \cdot \frac{s}{\sqrt{n}} \qquad (1)$$

where $t_{(1-\alpha/2;\, n-1)} \cdot \frac{s}{\sqrt{n}}$ is termed the *margin of error,* $\bar{x}$ is the mean of *n* sample observations, with a standard deviation, s, all of which have the same dimensional units. The right side critical value, $t_{(1-\alpha/2;\, n-1)}$ is the value that serves as the boundary between sample statistics that are likely to occur from those that are unlikely to occur for the Student t-distribution with n-1, degrees of freedom for a confidence level percentile of 100(1-α)%. In other words, the critical value is the lower bound limit of integration (the upper bound limit being infinity, ∞) for the improper integral of the probability density function (pdf) of the Student t-distribution having an evaluation of α/2. Alternately, it is the upper bound limit of integration (the lower bound limit being negative infinity, -∞) of the improper integral with the pdf as the integrand, yielding a cumulative distribution function (cdf) with a value of 1-α/2. For the most common confidence level percentiles, the right side critical value can be read directly as found in tables of critical values for the Student t-distribution for various degrees of freedom (Appendix C), or computed as shown in Appendix D. Due to the symmetry of the Student t-distribution, only the right side critical value need be listed in Appendix C or the (1-α/2) inverse cdf, computed in Appendix D.

For the 10 frankfurter observations, *n* = 10; significance level, α = 0.05; the 95% confidence interval using equation (1) is

$$\left[\mu_{lower}, \mu_{upper}\right] = \bar{x} \pm t_{0.975;\, 9} \cdot \frac{s}{\sqrt{n}} = 21.90 \pm 2.262 \cdot \frac{4.134}{\sqrt{10}} = 21.9 \pm 3.0 = [18.9, 24.9]$$

That is, the 95% confidence interval for the estimate of the population mean of the proportion of frankfurter fat is 21.9 ± 3.0%.

Thus we conclude that if the sampling of size *n*, should continue many times, we are 95% confident, that the sample means would be contained within the interval of 18.9 and 24.9% fat.

Prediction Intervals

For a two-sided 100(1-α)% prediction interval (P.I.), to contain the mean of *m* future observations, independent and randomly selected described by a normal distribution is

$$[\bar{y}_{lower}, \bar{y}_{upper}] = \bar{x} \pm t_{(1-\alpha/2;\, n-1)} \cdot s\sqrt{\frac{1}{m}+\frac{1}{n}} \qquad (2)$$

where *m* is the future number of observations, $m = 1, 2, \cdots, n$.

A two-sided 100(1-α)% simultaneous prediction interval to contain the values of *all m* future randomly selected observations from a previously sampled population, that can be described by a normal distribution, can be calculated using either the normal distribution or the Student t-distribution approach. Using the Student t-distribution the prediction interval is

$$[y_{lower}, y_{upper}] = \bar{x} \pm t_{(1-\alpha/(2m));\, n-1)} \cdot s\sqrt{\frac{1}{m}+\frac{1}{n}} \qquad (3)$$

where *simultaneous* indicates we are concerned with simultaneously containing all of the *m* observations within the calculated confidence interval with the selected level of confidence (note the difference in the lower percentile notation for the *test-statistic* between Equation 2 and Equation 3). In the case where *m* = 1 Equation 2 and Equations 3 are identical and the terms *mean* and *simultaneous* are extraneous. Therefore, where only a single future observation (*m* = 1) is made and using the frankfurter example, equation 3 yields

$$[y_{lower}, y_{upper}] = \bar{x} \pm t_{(1-\alpha/(2m));\, n-1)} \cdot s\sqrt{1+\frac{1}{n}} = 21.9 \pm 9.9 = [12.1, 31.7]$$

The 95% prediction interval is 21.9 ± 9.8%, or 12.1 to 31.7%. We interpret this result as follows: if in the future a single frankfurter is observed from the same population (assumed to be normally distributed) as that sampled, there is a 95% probability that the proportion of fat contained in the frankfurter would lie between 12.1 and 31.7%.

Prediction interval calculations using the normal distribution approach, although not shown here, indicate prediction interval bounds similar to those calculated using the Student t-distribution when *n* > 10 and *m* = 1. However, for cases where *n* < 10 or *m* > 1 the percent difference of the prediction interval bounds between the two approaches becomes more pronounced. Under these conditions the Student t-distribution approach provides more conservative prediction interval bounds than that of the normal distribution approach. In terms of the prediction interval lower bound, the student t-distribution approach provides a lesser (more conservative) value than the normal distribution approach.  Providing a more conservative prediction interval lower bound through use of the Student t-distribution approach lends itself well to more serious-minded applications like that found in the following section.  Figure 1. compares the two-sided prediction interval factors for the normal and Student-t distributions for a family of curves for *m* future observations at several values of *n* for α = 0.05. In each case the Student-t distribution provides a more conservative (wider) prediction interval.

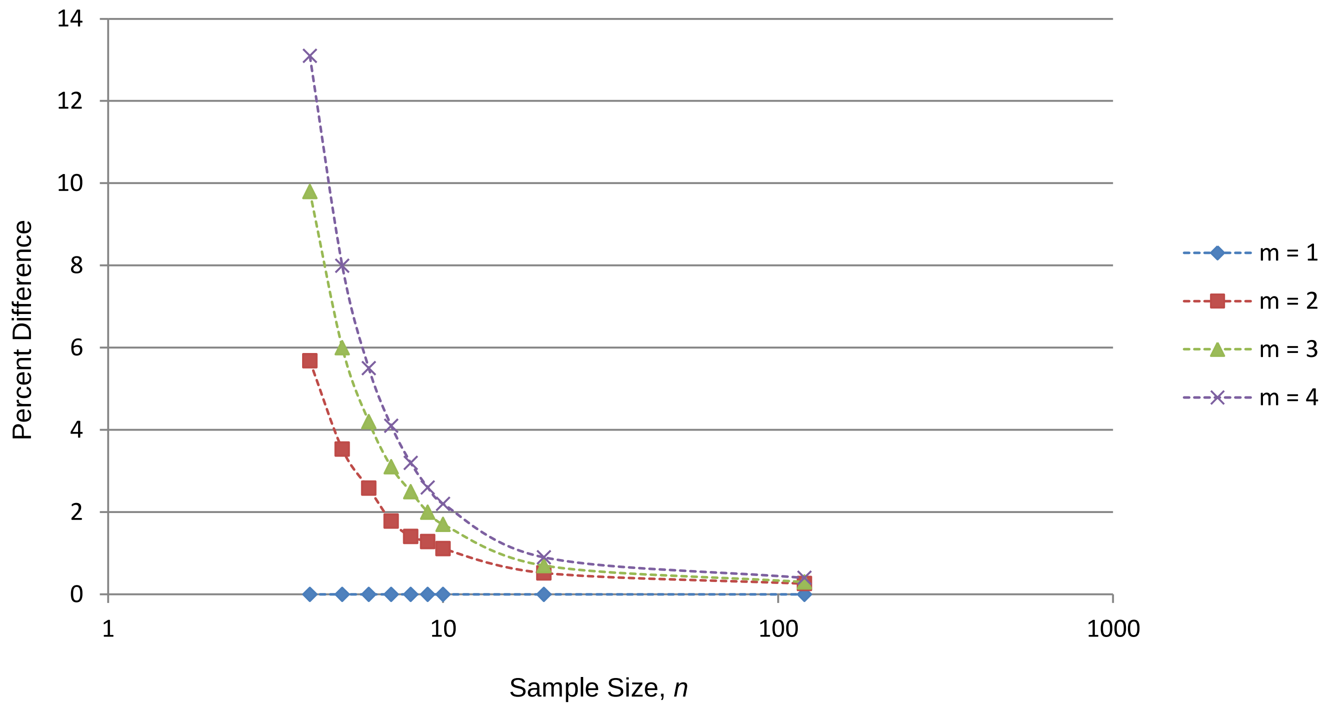



**Figure 1.** Comparison between normal and Student-t distributions two-sided prediction interval factors versus sample size for α = 0.05

## APPLICATION

<u>Canister Durations</u>

The following duration data (in minutes) were obtained from NEDU’s tests of canisters for a particular UBA tested at a single water temperature:

121.6, 114.6, 118.9, 122.0, 114.8, 116.2, 106.0, 113.4, 109.7, and 112.8

This single sample consisting of ten observations, represent the elapsed time from the start of $CO_2$ inflow into the scrubber canister until the $CO_2$ percentage in the canister effluent reached 0.5%.

The mean for this data was 115.0 min, and the standard deviation was 5.0 min. Assuming a normally distributed population, from Equation 1 the 95% *C.I.* for the mean was 111.4 to 118.6 min. From Equation 2, the 95% *P.I.* was 103.1 to 126.9 min.

That is, if the sampling were repeated many times with the same sample size of 10, from the same canister population tested under identical conditions, we anticipate that 97.5% of the time the sample *mean* would be greater than 111 min. It would not be likely to exceed 119 min.

However, if a diver were, in essence, to make an observation during his next dive, then we predict that his canister would be likely to last longer than 103 min. The chance that his canister duration would be less than 103 min is essentially only 2.5%.

With our focus on diver safety, we are concerned about the chances that a diver's canister will last less than the published canister duration — the duration he is expecting. We are thus interested in *m* future prediction interval bounds (Equations 2 and 3). We have little interest in the mean or average duration of the canister under the same dive conditions and thus are less concerned about the confidence bounds on the mean (Equation 1).

<u>Monte Carlo Study</u>

One of the uses for Monte Carlo simulation techniques is in providing an intuitive interpretation of statistical equations. We tested the equation for one-sample prediction interval bounds in just that manner, obtaining samples of various sizes from large populations meeting prescribed statistical distributions. We used Mathcad (version 11, MathSoft; Cambridge, MA) to generate random numbers distributed normally (that is, in a Gaussian fashion) with a target mean of 115 min and a standard deviation of 5.0 min — figures matching the canister duration data determined in NEDU's experiments.

<u>*N = 50,000*</u>

In our first sampling, 50,000 normally distributed pseudorandom data points were generated (Appendix D). The mean for that large sample was 115.025, and the standard deviation was 5.001. Due to the large sample size, the estimate on the population mean was tight: the 95% *C.I.* on the mean ranged from 114.98 to 115.07 min. Using Equation 2, the 95% prediction interval bounds ranged from 105.22 to 124.83 min.

Based on the definition of the lower 95% prediction interval bound, 2.5% of the durations should have been less than 105.22 min. In fact, that particular sample had 1289 such durations, comprising 2.6% of the total data. Thus, in this case the lower prediction interval bound defined by Equation 2 accurately predicted a canister duration that would be exceeded 97.5% of the time, and would *not* be exceeded only about 2.5% of the time.

<u>*N = 100*</u>

In our next sample, 100 normally distributed pseudorandom data points were generated. For a particular sampling, the sample mean was 115.51 min, and the standard deviation was 5.06 min. The 95% *C.I.* for the true population mean ranged from 114.51 to 116.51 min — not as tight as in the first example, but nevertheless narrow and centered on the sample mean.

From Equation 2, the 95% prediction interval bounds ranged from 105.43 to 125.59 min. Although the sample size of 100 was 500 times smaller than in the first example, the estimate for the lower 95% prediction interval bound varied only slightly — from 105.22 min in the first instance to 105.43 min in the smaller sample.

When the entire data set was sorted, only two "durations" were in fact less than 105.43 minutes and returned a fraction of 0.02, as close as possible to the ideal fraction of 0.025 for a sample size of 100.

Inverse Cumulative Distribution Function

A third method of estimating the lower 95% prediction interval bound can be found directly from the inverse cumulative probability distribution function (iCDF) for the normal population. The iCDF takes a probability $(\alpha/2)$ as an argument and returns a value (equal to or less than) for the random variable (assumed to be normally distributed) whose probability is the argument.

The usefulness of the iCDF for normally distributed data is that it is not dependent on either an actual dataset or a dataset simulated by a Monte Carlo technique. It is based simply on the mathematics of the normal distribution. Since this function can be found from numerous sources, including statistical tables, we will not describe it in detail. In Appendix D, the iCDF is represented by qnorm$\left(\alpha/2, \bar{Y}, s\right)$.

Table 1 shows the result of multiple Monte Carlo simulations with the number of observations in each sample ranging from 3 to 50,000. The last column shows the iCDF for the lower 2.5% value for a distribution with mean and *s* matching each dataset.

Table 1. Analysis of multiple datasets produced by Monte Carlo methods.

| n | $\bar{Y}$ | $s$ | $CL_{low}$ | $CL_{up}$ | $PI_{low}$ | $PI_{up}$ | iCDF |
|---|---|---|---|---|---|---|---|
| 50,000 | 115.02 | 5.01 | 114.98 | 115.07 | 105.22 | 124.83 | 105.22 |
| 100 | 115.14 | 4.98 | 114.15 | 116.13 | 105.21 | 125.07 | 105.38 |
| 30 | 114.65 | 4.91 | 112.81 | 116.48 | 104.43 | 124.86 | 105.21 |
| 20 | 114.97 | 5.04 | 112.61 | 117.32 | 104.17 | 125.77 | 105.10 |
| 10 | 115.17 | 5.05 | 111.56 | 118.78 | 103.19 | 127.14 | 105.27 |
| 5 | 115.15 | 4.89 | 109.08 | 121.23 | 100.27 | 130.04 | 105.56 |
| 3 | 114.71 | 4.62 | 103.22 | 126.19 | 91.74 | 137.67 | 105.65 |

$CL_{low}$ = lower bound of the 95% confidence interval, $CL_{up}$ = upper bound of the 95% confidence interval, $PI_{low}$ = lower bound of the 95% prediction interval, $PI_{up}$ = upper bound of the 95% prediction interval, iCDF = inverse cumulative distribution function.

In the first Monte Carlo sampling (n = 50,000), for a probability of 0.025, a mean of 115.02 min, and a standard deviation of 5.01 min, the iCDF yields 105.22 min. That precisely matches the lower 95% prediction limit found from Equation 2.

For the second sample with n = 100, the iCDF of the normal population for a probability of 0.025, a mean of 115.14 min, and a standard deviation of 4.98 min returned a value of 105.38 min. That is reasonably close to the lower 95% prediction limit of 105.21 min.

Figure 2 compares the values in the Table 1 iCDF column with the 100(1-α)% lower bound value for the prediction interval $PI_{low}$ column using the percent difference calculated as the quotient of the difference between the values divided by their mean. The percent difference between these values ranges from 0% with an n of 50,000 to 0.9% with an n of 20. In other words, the estimated lower prediction interval bound is very close (within 1%) to what is expected on the basis of the properties of the normal population. For a sample size of 5, the percent difference is about 5%.

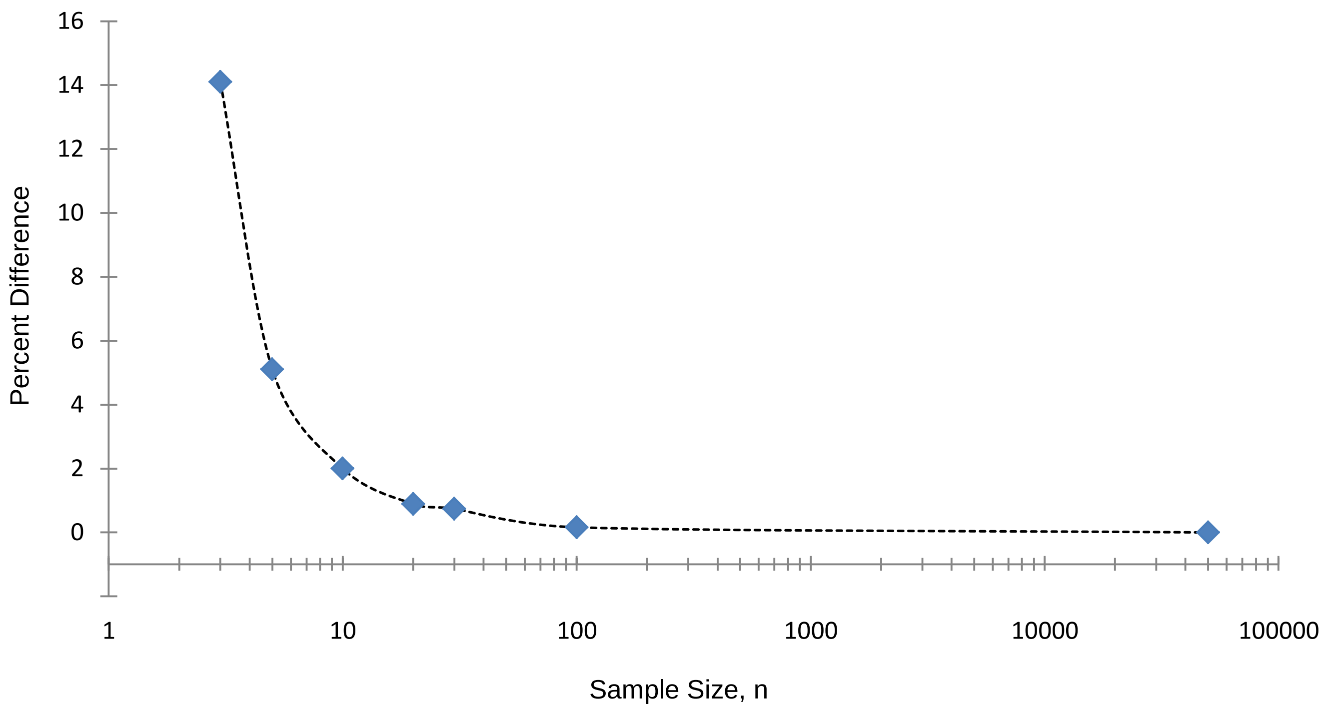



**Figure 2.** Comparison between iCDF and $PI_{low}$.

## CONCLUSIONS

As described in engineering statistics texts, the use of one-sample prediction intervals is profitably applied to the testing of closed-circuit UBA canister durations. Their use in “quick-look” evaluations of UBA complements their use in the complete UBA testing procedures described in NEDU TM 01-12 and Tech Report 2-99.

Calculating the iCDF for the normal distribution is mathematically complex requiring an understating of definite integral calculus. However, without employing the rigors of the calculus, readily available statistics and mathematical software packages make finding approximations to those values straightforward. The ease of making those computations makes it possible to predict a canister duration limit on the basis of an assumed mean canister duration and an assumed standard deviation. It can also be used to compare prediction interval bounds based on Equation 2. For sample sizes of 10 or more, the resulting values should agree within a few percentage points.

## REFERENCES


1. J. R. Clarke, *Statistically Based $CO_2$ Canister Duration Limits for Closed-Circuit Underwater Breathing Apparatus*, NEDU TR 2-99, Navy Experimental Diving Unit, April 1999.

2. J. Clarke, K. Russel, and L. Crepeau, *MK 16 Canister Limits for SDV Operations*, NEDU TR 2-93, Navy Experimental Diving Unit, June 1993.

3. L. Crepeau and J. Clarke, *Recommended Canister Limits for the Draeger LAR V/MK 25 UBA Using 408 L-Grade and 812 D-Grade Sofnolime*, NEDU TR 9-97, Navy Experimental Diving Unit, June 1998.

4. J. L. Devore, *Probability and Statistics for Engineering and the Sciences*, $5^{th}$ ed. (Pacific Grove, CA: Duxbury Press, 2000), pp. 299–301.

5. R. L. Scheaffer and J. T. McClave, *Probability and Statistics*, $3^{rd}$ ed. (Boston, MA: PWS-Kent, 1990), pp. 291–292.

6. G. G. Vining, *Statistical Methods for Engineers* (Pacific Grove, CA: Duxbury, 1998), p. 176.

7. S. L. Beilken, L. M. Eadie, P. N. Jones, and P.V. Harris, “Sensory and Mechanical Assessment of the Quality of Frankfurters,” *J. Texture Studies*, Vol. 21 (1990), pp. 395–409.

8. H. Vaghadia, M. A. Solylo, C. L. Henderson, and G. W. Mitchell, **“**Selective Spinal Anesthesia for Outpatient Laparoscopy. II: Epinephrine and Spinal Cord Function,” *Can J Anaesth.*, Vol. 48, No. 3 (March 2001), pp. 261–266.

9. G. Hahn and W. Meeker, *Statistical Intervals* (New York: Wiley, 1991).

## Appendix A: Prediction Limits on Regression

Some curve fitting software can plot not only the best estimate for fitted data but also confidence intervals on the best fit and prediction limits for the data. One such software package is Systat's Table Curve 2D (Systat Software Inc.; Richmond, CA), formerly made by Jandel Scientific (San Rafael, CA).

For regressions, prediction limits are found from the following equations:[1]

$$P.I. = \hat{Y}_h \pm t(1-\alpha/2, dof) \cdot s\{Y_{h(new)}\} ,$$

where

$$s\{Y_{h(new)}\} = \sqrt{MSE} \cdot \sqrt{\left[1 + \frac{1}{n} + \frac{\left(X_h - \bar{X}\right)^2}{\sum\left(X_i - \bar{X}\right)^2}\right]},$$

which is the estimated standard deviation of the new predicted Y.

The square root of the mean square error, $\sqrt{MSE}$ is the standard error of estimate (fit standard error), $s_e$, and is calculated using:

$$s_e = \sqrt{\frac{\sum(y-\hat{y})^2}{n-2}} = \sqrt{\frac{\sum y^2 - b_0 \sum y - b_1 \sum xy}{n-2}}$$

where the constant coefficients $b_1$ and $b_0$, are the slope and y-coordinate of the y-intercept of the linear regression equation respectively, all summations iterate from 1 to n, with n being the number of bivariate pairs in the data set and the degrees of freedom are n-2.

Since $s_e = \sqrt{MSE}$ , the fit standard error (root mean square error)

$$s\{Y_{h(new)}\} = s_e \cdot \sqrt{\left[1 + \frac{1}{n} + \frac{\left(X_h - \bar{X}\right)^2}{\sum\left(X_i - \bar{X}\right)^2}\right]}.$$

Therefore,

$$P.I. = \hat{Y}_h \pm t(1-\alpha/2, dof) \cdot s_e \cdot \sqrt{\left[1 + \frac{1}{n} + \frac{\left(X_h - \bar{X}\right)^2}{\sum\left(X_i - \bar{X}\right)^2}\right]}$$

Figure A-1, from NEDU TR 2-99, shows 1000 simulated canister durations distributed in a Gaussian fashion around mean durations. In this simulation the mean durations were curvilinearly dependent upon temperature, much as they are in the MK 25 UBA.

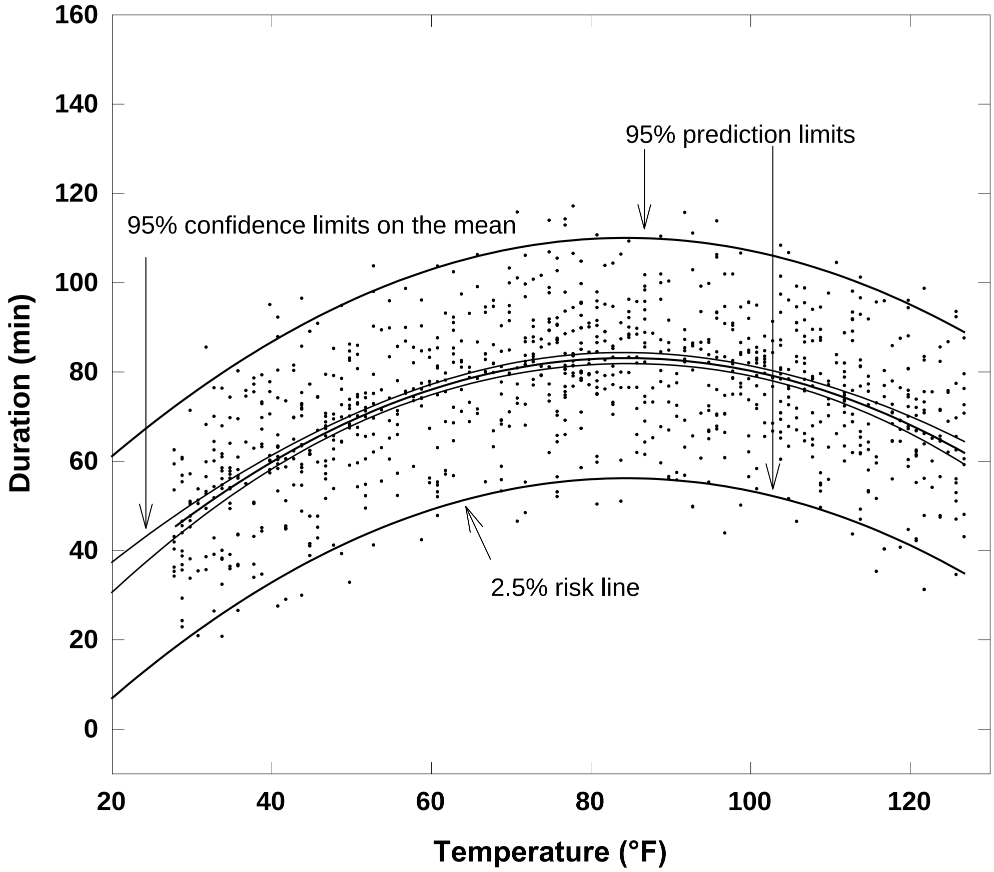



**Figure A-1.** 1000 simulated, normally distributed canister durations with curvilinear temperature dependencies.

The innermost line is the best estimate for the mean duration as a function of temperature. That line is surrounded closely by the 95% confidence limits on the mean. The furthermost curved lines mark the boundary for the 95% prediction interval. As expected, about 2.5% of the data lies above the upper prediction limit line, and about 2.5% lies below the lower prediction limit.

**Table Curve Equations**

Sum of Squares due to Error: $SSE = \sum \omega_i \cdot \left(y_i - \hat{y}_i\right)^2$ ,

where $\omega_i$ = weight, $y_i$ = y data value, $\hat{y}_i$ = predicted y value, and n = number of data points.

Mean Square Error: $MSE = \dfrac{SSE}{DOF}$ .

Degree of freedom: $DOF = n - m$,
where $m$ = number of coefficients (parameters) fitted.

Fit Standard Error (root MSE): $S_e = \sqrt{MSE}$ .

$$P.I. = \hat{y} \pm t \cdot \sqrt{MSE} \cdot \sqrt{1 + l' \cdot \left( (X'X)^{-1} \right) \cdot l}$$

Or

$$P.I. = \hat{y} \pm t \cdot s_e \cdot \sqrt{1 + l' \cdot \left( (X'X)^{-1} \right) \cdot l}$$

where $t$ = Student's t for given confidence level and *DOF*, $l$ = coefficient partial derivative vector evaluated at $x_i$, and $\left( (X'X)^{-1} \right)$ = inverse of the design matrix typically found in regressions using matrices.[2]

Compare this to

$$P.I. = \hat{Y}_h \pm t \cdot s_e \cdot \sqrt{\left[ 1 + \frac{1}{n} + \frac{\left( X_h - \overline{X} \right)^2}{\sum \left( X_i - \overline{X} \right)^2} \right]}.$$

To see how single-sample prediction limits relate to regression-based prediction limits, the canister durations (shown on page 3) obtained during actual NEDU measurements were replicated and assigned to three equally spaced temperatures. Since the data sets at each temperature were identical, a linear model was applied to the data. TableCurve 2D, Version 4.0 was used for the fitting. Figure B-2 shows the results.

Replication of the original ten data points to form thirty data points resulted in a lower 95% prediction limit ranging from 104.2 to 104.5 min. If we had published a canister limit based on that data, it would have been only about one minute longer than the limit chosen with the single sample. The resulting published durations would differ by about 1%.

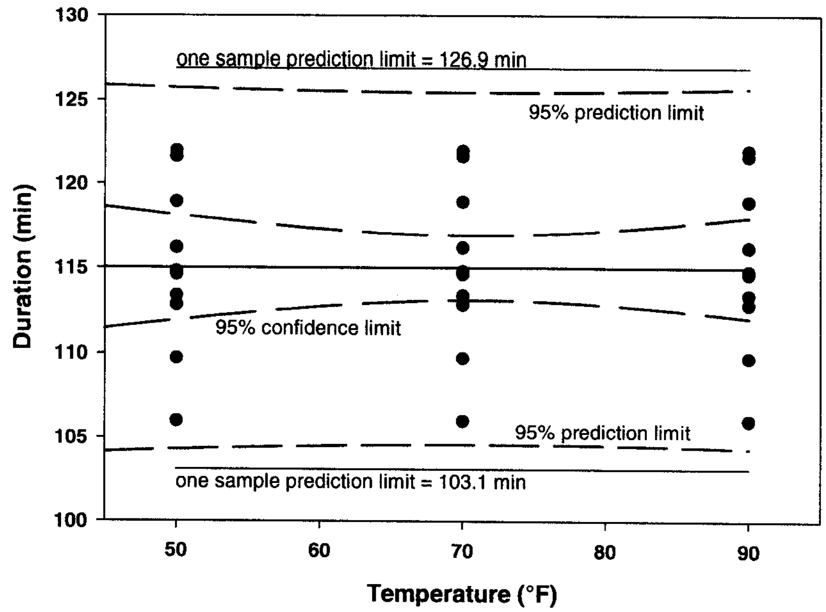



**Figure A-2.** Prediction limits based on regression, for comparison to a single-sample prediction limit.

## Appendix A References


1. J. Neter, W. Wasserman, and M. H. Kutner, *Applied Linear Statistical Models* (Homewood, IL: Richard D. Irwin, Inc., 1990), p. 82.

2. *Ibid.*, p. 206.

## Appendix B: Textbook Examples

1. Scheaffer and McClave (1990), p. 292.

Ten independent observations are taken on bottles coming from a machine designed to fill them to 16 ounces. The n = 10 observations show a mean of 16.1 oz and a standard deviation of 0.01 oz. Find a 95% prediction interval for the ounces of fill in the next bottle to be observed.

Solution:

$$\bar{\chi} \pm t_{0.025}\, s\sqrt{1+\frac{1}{n}}$$

$$16.1 \pm (2.262)(0.01)\sqrt{1+\frac{1}{10}}\,.$$

With $t_{0.025}$ based on 9 degrees of freedom, we get

16.1 ± 0.024, or (16.076, 16.124).

We are 95% confident that the next observation will lie between 16.076 and 16.124.

2. Scheaffer and McClave (1990), p. 293.

A particular subcompact automobile has been tested for gas mileage 50 times. These mileage figures have a mean of 39.4 mpg and a standard deviation of 2.6 mpg. Predict the gas mileage to be obtained on the next test, with $1-\alpha = 0.90$.

Answer: *P.I.* = 34.998, 43.802; *C.I.* = 38.784, 40.016

3. Scheaffer and McClave (1990), p. 293.

It is extremely important for a business firm to be able to predict the amount of downtime for its computer system over the next month. A study of the past five months has shown the downtime to have a mean of 42 hours and a standard deviation of 3 hours. Predict the downtime for next month in a 95% prediction interval.

Answer: *P.I.* = 32.877, 51.123; *C.I.* = 38.275, 45.725

4. Scheaffer and McClave (1990), p. 292.

In studying the properties of a particular resistor, the actual resistances produced were measured on a sample of 15 resistors. These resistances had a mean of 9.8 ohms and a standard deviation of 0.5 ohm. One resistor of this type is to be used in a circuit. Find a 95% prediction interval for the resistance it will produce.

Answer: *P.I.* = 8.69, 10.91; *C.I.* = 9.523, 10.077

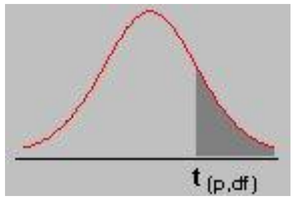
5. Vining (1998), p. 177.

A company manufactures high-voltage power supplies with a nominal output of 350 volts. The production people are concerned that the process is beginning to produce power supplies with a true mean output voltage somewhat greater than the nominal value. The last four power supplies tested at 351.4, 351.5, 351.2, and 351.6 volts. Conduct the most appropriate test at a 0.10 significance level, and construct a 90% confidence interval for the true mean voltage and 90% prediction intervals for the voltages.

$H_a$: $\mu > 350$, t = 16.69, reject $H_{0.}$ *C.I.* = 351.224, 351.626; *P.I.* = 350.976, 351.874

# Appendix C: Upper Critical Values of the Student's T-Distribution

From http://www.itl.nist.gov/div898/handbook/eda/section3/eda3672.htm




| df\p | 0.40 | 0.25 | 0.10 | 0.05 | 0.025 | 0.01 | 0.005 | 0.0005 |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.324920 | 1.000000 | 3.077684 | 6.313752 | 12.70620 | 31.82052 | 63.65674 | 636.6192 |
| 2 | 0.288675 | 0.816497 | 1.885618 | 2.919986 | 4.30265 | 6.96456 | 9.92484 | 31.5991 |
| 3 | 0.276671 | 0.764892 | 1.637744 | 2.353363 | 3.18245 | 4.54070 | 5.84091 | 12.9240 |
| 4 | 0.270722 | 0.740697 | 1.533206 | 2.131847 | 2.77645 | 3.74695 | 4.60409 | 8.6103 |
| 5 | 0.267181 | 0.726687 | 1.475884 | 2.015048 | 2.57058 | 3.36493 | 4.03214 | 6.8688 |
| | | | | | | | | |
| 6 | 0.264835 | 0.717558 | 1.439756 | 1.943180 | 2.44691 | 3.14267 | 3.70743 | 5.9588 |
| 7 | 0.263167 | 0.711142 | 1.414924 | 1.894579 | 2.36462 | 2.99795 | 3.49948 | 5.4079 |
| 8 | 0.261921 | 0.706387 | 1.396815 | 1.859548 | 2.30600 | 2.89646 | 3.35539 | 5.0413 |
| 9 | 0.260955 | 0.702722 | 1.383029 | 1.833113 | 2.26216 | 2.82144 | 3.24984 | 4.7809 |
| 10 | 0.260185 | 0.699812 | 1.372184 | 1.812461 | 2.22814 | 2.76377 | 3.16927 | 4.5869 |
| | | | | | | | | |
| 11 | 0.259556 | 0.697445 | 1.363430 | 1.795885 | 2.20099 | 2.71808 | 3.10581 | 4.4370 |
| 12 | 0.259033 | 0.695483 | 1.356217 | 1.782288 | 2.17881 | 2.68100 | 3.05454 | 4.3178 |
| 13 | 0.258591 | 0.693829 | 1.350171 | 1.770933 | 2.16037 | 2.65031 | 3.01228 | 4.2208 |
| 14 | 0.258213 | 0.692417 | 1.345030 | 1.761310 | 2.14479 | 2.62449 | 2.97684 | 4.1405 |
| 15 | 0.257885 | 0.691197 | 1.340606 | 1.753050 | 2.13145 | 2.60248 | 2.94671 | 4.0728 |
| | | | | | | | | |
| 16 | 0.257599 | 0.690132 | 1.336757 | 1.745884 | 2.11991 | 2.58349 | 2.92078 | 4.0150 |
| 17 | 0.257347 | 0.689195 | 1.333379 | 1.739607 | 2.10982 | 2.56693 | 2.89823 | 3.9651 |
| 18 | 0.257123 | 0.688364 | 1.330391 | 1.734064 | 2.10092 | 2.55238 | 2.87844 | 3.9216 |
| 19 | 0.256923 | 0.687621 | 1.327728 | 1.729133 | 2.09302 | 2.53948 | 2.86093 | 3.8834 |
| 20 | 0.256743 | 0.686954 | 1.325341 | 1.724718 | 2.08596 | 2.52798 | 2.84534 | 3.8495 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **21** | 0.256580 | 0.686352 | 1.323188 | 1.720743 | 2.07961 | 2.51765 | 2.83136 | 3.8193 |
| **22** | 0.256432 | 0.685805 | 1.321237 | 1.717144 | 2.07387 | 2.50832 | 2.81876 | 3.7921 |
| **23** | 0.256297 | 0.685306 | 1.319460 | 1.713872 | 2.06866 | 2.49987 | 2.80734 | 3.7676 |
| **24** | 0.256173 | 0.684850 | 1.317836 | 1.710882 | 2.06390 | 2.49216 | 2.79694 | 3.7454 |
| **25** | 0.256060 | 0.684430 | 1.316345 | 1.708141 | 2.05954 | 2.48511 | 2.78744 | 3.7251 |
| | | | | | | | | |
| **26** | 0.255955 | 0.684043 | 1.314972 | 1.705618 | 2.05553 | 2.47863 | 2.77871 | 3.7066 |
| **27** | 0.255858 | 0.683685 | 1.313703 | 1.703288 | 2.05183 | 2.47266 | 2.77068 | 3.6896 |
| **28** | 0.255768 | 0.683353 | 1.312527 | 1.701131 | 2.04841 | 2.46714 | 2.76326 | 3.6739 |
| **29** | 0.255684 | 0.683044 | 1.311434 | 1.699127 | 2.04523 | 2.46202 | 2.75639 | 3.6594 |
| **30** | 0.255605 | 0.682756 | 1.310415 | 1.697261 | 2.04227 | 2.45726 | 2.75000 | 3.6460 |
| | | | | | | | | |
| **inf** | 0.253347 | 0.674490 | 1.281552 | 1.644854 | 1.95996 | 2.32635 | 2.57583 | 3.2905 |

# Appendix D: Mathcad Implementation of Monte Carlo Analysis

$n := 50000$ $\qquad\qquad$ $A := rnorm(n, 115, 5)$

$mean(A) = 115.025$

Sample standard deviation:

$$s := \sqrt{\frac{n}{n-1} \cdot var(A)} \qquad\qquad s = 5.001$$

Degrees of freedom:

$$df := n - 1 \qquad\qquad df = 5 \times 10^{4}$$

Enter level of significance: $\alpha := 0.05$ $\qquad$ Confidence level: $1 - \alpha = 95\%$

<u>Confidence Limit determination:</u>

critical value:

$$t_0 := qt\left(1 - \frac{\alpha}{2}, df\right) \qquad\qquad t_0 = 1.96$$

upper limit:

$$U := mean(A) + t_0 \cdot \frac{s}{\sqrt{n}} \qquad\qquad U = 115.069$$

lower limit:

$$L := mean(A) - t_0 \cdot \frac{s}{\sqrt{n}} \qquad\qquad L = 114.981$$

<u>Prediction Limit determination:</u>

$$PIU := mean(A) + t_0 \cdot s \cdot \sqrt{1 + \frac{1}{n}} \qquad\qquad PIU = 124.827$$

$$PIL := mean(A) - t_0 \cdot s \cdot \sqrt{1 + \frac{1}{n}} \qquad\qquad PIL = 105.223$$

$i := 0..n-1$ $\qquad$ $C_i := \text{if}\left(A_i < PIL, 1, 0\right)$

$shorts := \sum C$ $\qquad$ $shorts = 1.289 \times 10^3$

$fraction := \frac{shorts}{n}$ $\qquad$ $fraction = 0.026$

$c := \text{qnorm}(0.025, \text{mean}(A), s)$ $\qquad$ $c = 105.223$

$\text{qnorm}(0.05, \text{mean}(A), s) = 106.799$